

Gérer la complexité de l'écosystème numérique actuel

L'Allemagne, à travers le *Deutsches Institut für Normung* (DIN), joue un rôle central dans l'élaboration et la promotion de normes internationales, facilitant ainsi l'interopérabilité et la confiance dans le secteur de l'information et de la documentation. Deux exemples récents illustrent cette contribution.



L'INTERNATIONAL STANDARD CONTENT CODE (ISCC) : UNE NOUVELLE NORME POUR L'IDENTIFICATION DES CONTENUS

L'expansion rapide des contenus numériques au cours de la dernière décennie a transformé notre manière de créer, gérer et partager l'information. Le volume considérable d'actifs numériques produits quotidiennement a introduit de nouveaux défis en matière d'organisation, de recherche et de préservation des connaissances. Bien que les identifiants traditionnels – comme les ISBN ou les DOI – restent essentiels pour le catalogage d'œuvres abstraites, ils n'ont pas été conçus pour gérer la complexité de l'écosystème numérique actuel.

L'*International Standard Content Code* (ISCC), formalisé sous ISO 24138:2024, comble cette lacune en liant les identifiants existants aux contenus de manière unique et reproductible via des algorithmes *open source*, favorisant ainsi l'interopérabilité et permettant aux systèmes de catalogage traditionnels d'évoluer parallèlement à des méthodes d'identification dérivées du contenu.

Envie de découvrir le fonctionnement de l'ISCC ? Testez notre démonstrateur *open source* sur ISCC Playground : un outil interactif pour générer et comparer des codes à partir d'actifs numériques, et explorer concrètement ses usages.

Au cœur de l'ISCC se trouve un identifiant multi-composants, conservant les similarités, et agissant comme une empreinte du contenu numérique. Contrairement aux systèmes traditionnels qui attribuent des identifiants via une inscription externe, l'ISCC est généré directement à partir du contenu numérique lui-même. Cela signifie que tout contenu – texte, image, audio ou vidéo – peut être identifié et comparé via des algorithmes *open source*.

Par exemple, un code ISCC réel pourrait ressembler à ceci :

ISCC:KAA2Y5NUST7BFD5NN2XIDK7VW3WG4OEPMRQNPk37TE

L'ISCC se compose de plusieurs unités, chacune capturant différents aspects du contenu numérique pour une identification précise :

- **Meta-Code**: encode la similarité syntaxique et lexicale des métadonnées.

- **Semantic-Code**¹: capture la similarité conceptuelle et sémantique du contenu.

- **Content-Codes (Texte, Image, Audio, Vidéo)**: encodent la similarité perceptuelle, syntaxique et structurelle.

- **Data-Code**: encode la similarité des données brutes.

- **Instance-Code**: identifie des instances spécifiques de données, similaire à un *checksum* ou hachage cryptographique.

Cette structure permet à différents systèmes de collaborer efficacement, en facilitant la détection des doublons, la synchronisation des enregistrements et la traçabilité des origines de contenu. Même lorsqu'un contenu est modifié ou reformaté, l'ISCC permet de l'identifier, ce qui en fait un outil précieux pour la gestion et la préservation des contenus numériques.

Le développement de l'ISCC n'a pas été sans défis. Le premier obstacle ? Convaincre les parties prenantes des systèmes d'identification traditionnels que l'ISCC vise à compléter, et non à remplacer, leur travail. En toute logique, certains se sont tout d'abord montrés sceptiques – il est difficile de remettre en question des systèmes bien établis. Toutefois, grâce à des discussions continues, le rôle complémentaire de l'ISCC a été clarifié.

Un autre défi était le fait que l'ISCC ne nécessite aucune autorité centrale d'enregistrement. Pour ceux habitués aux identifiants traditionnels, un identifiant indépendant semblait un concept étranger. Mais puisque l'ISCC est généré directement à partir de l'actif numérique, il n'y a aucun besoin d'attribution manuelle – n'importe qui peut dériver le même ISCC à partir du même contenu.

Le développement de l'ISCC a permis de tirer plusieurs enseignements clés :

- **L'interopérabilité est essentielle** : un identifiant distribué et dérivé du contenu permet une communication fluide entre plateformes et secteurs.

- **L'*open source* accélère les progrès** : la transparence et la collaboration ont favorisé une

innovation rapide et une large validation de l'ISCC.

- **Faire le lien entre données et sens est un défi permanent** : l'ISCC jette les bases pour de futures améliorations, notamment l'intégration des codes sémantiques pour mieux saisir le contenu conceptuel.

Ces leçons ont non seulement façonné l'ISCC, mais ont aussi apporté des orientations utiles à des efforts plus larges de normalisation dans la gestion du contenu numérique.

Dès le départ, la vision de l'ISCC était ambitieuse : créer un identifiant universel applicable à tous les formats numériques, sans dépendance à des registres centraux. Les objectifs initiaux visaient à rendre le système robuste, interopérable et adaptable aux évolutions technologiques futures.

Aujourd'hui, l'ISCC est passé de la théorie à la pratique. Il est reconnu pour sa capacité à simplifier l'identification de contenus dans divers secteurs, des bibliothèques universitaires et institutions de recherche aux médias et archives culturelles. Toutefois, son succès à long terme repose sur une collaboration continue et une adoption à grande échelle.

L'Allemagne est depuis longtemps reconnue pour la précision et la fiabilité de ses normes, – des organismes comme le DIN jouant un rôle majeur dans les normes mondiales. Le développement de l'ISCC s'inscrit dans cette tradition, en intégrant des solutions de pointe fondées sur les données aux pratiques établies en matière d'identification.

En associant l'ISCC à l'approche stratégique du DIN, les parties prenantes du monde académique, culturel et industriel peuvent s'appuyer sur un système capable de répondre aux défis actuels tout en restant évolutif pour l'avenir.

TITUSZ PAN,
ISCC Foundation
tp@iscc.io

[1] Bien que la norme ISO 24138:2024 réserve un type-ID pour les « *Semantic Codes* », les algorithmes spécifiques ne sont pas encore définis.

LE GLOBAL MEDIA IDENTIFIER¹: RÉTABLIR LA CONFIANCE DANS L'ÉCOSYSTÈME MÉDIATIQUE

En 2016, une femme de 53 ans, originaire de l'État de Saxe, en Allemagne de l'Est, a payé 294 euros pour un billet d'avion qu'elle n'a jamais utilisé. Elle pensait partir à Porto, au Portugal, alors que le billet fourni par son agence de voyage la conduisait à Bordeaux, en France. Dans son accent local, les noms des deux villes se prononcent presque de la même manière. Apparemment, chaque année, certaines personnes atterrissent à Rodez, en France, pour se rendre compte, à leurs dépens, que ce n'est pas Rhodes, en Grèce, leur destination prévue. Sydney, en Nouvelle-Écosse (Canada), accueille également régulièrement des voyageurs égarés en route pour l'Australie.

Quand l'ambiguïté devient un problème, l'identification devient essentielle. Ce qui peut sembler technique – et ça l'est – peut non seulement éviter de mauvaises surprises aux vacanciers mal orientés, mais aussi protéger tout processus contre les erreurs, les confusions et les malentendus.

Un exemple connu est celui des codes d'aéroports, qui distinguent Portland dans le Maine (PWM), de Portland dans l'Oregon (PDX), ou encore de Portland en Australie (PTJ) – sans oublier Bordeaux (BOD) de Porto (OPO). D'autres exemples d'identifiants utiles dans notre quotidien incluent les ISBN pour les livres, ou les numéros d'identification fiscale. Une plaque d'immatriculation, un numéro de téléphone mobile ou encore une URL sont aussi des identifiants uniques.

L'attribution sans ambiguïté est dans l'intérêt de tous les acteurs économiques. Il n'est donc pas surprenant que les principes de type « clé du fournisseur » (KYS) et « clé du client » (KYC) soient devenus essentiels pour sécuriser les chaînes d'approvisionnement dans presque tous les secteurs. À l'inverse, l'absence d'attribution claire peut être exploitée par des acteurs malveillants qui usurpent l'identité de marques, fournisseurs ou acheteurs dans le but de tromper et d'escroquer – et de s'en tirer impunément. C'est précisément ce qui se passe aujourd'hui dans le domaine de la désinformation, à travers le phénomène des « *Doppelgängers* ». Il s'agit d'imposteurs anonymes se faisant passer – entre autres – pour *Der Spiegel*, Le Monde, la BBC ou CNN, afin de publier en ligne des contenus qui n'auraient jamais été diffusés par ces

médias reconnus. Dans certains cas, ces activités ont pu être retracées jusqu'à des entités étatiques engagées dans des opérations de manipulation et d'interférence de l'information étrangère (FIMI) – ou plus franchement, une forme de guerre hybride. Face à cette nouvelle forme de menace et à l'urgence d'y répondre, on réalise douloureusement qu'il n'existe aujourd'hui aucun protocole d'identification agréé, établi et lisible par machine dans la chaîne d'approvisionnement de l'information en ligne.

Un nouvel effort de normalisation par l'ISO vise à combler ce vide. L'organisation vient de publier un *International Workshop Agreement* (IWA), définissant la syntaxe et le modèle opérationnel d'un *Global Media Identifier* (GMI), sous la référence IWA44:2025, dans le but d'identifier clairement les sources de contenus en ligne. L'un des objectifs principaux du GMI est de réduire les risques d'erreurs ou de mauvaises attributions, qu'elles soient malveillantes (comme les *Doppelgängers*) ou simplement dues à la négligence ou à l'inattention.

Ce qui rend la proposition et l'implémentation de cet identifiant encore plus pertinente, c'est qu'il s'insère dans un écosystème déjà existant.

Premièrement, certains efforts liés au droit d'auteur visent à identifier les contenus – comme une image ou un texte – au nom de leurs auteurs, dans une logique de rémunération. Deuxièmement, la plupart des acteurs du secteur utilisent déjà leurs propres identifiants internes et propriétaires. En réalité, aucun grand jeu de données ne fonctionne sans une forme d'identification. Enfin, plusieurs systèmes d'identification voisins sont bien établis, comme ceux pour les entités légales (LEI) ou les créateurs de contenus (ISNI).

Le problème, cependant, est que ces systèmes ne sont pas interopérables. Le GMI ne vise donc pas à les remplacer, mais à les harmoniser en comblant le lien manquant. Cette approche prend d'autant plus de sens dans le contexte de l'essor rapide de l'intelligence artificielle, où il devient crucial de documenter et authentifier les contenus utilisés pour entraîner des modèles de langage. Cela s'inscrit aussi dans un cadre réglementaire en pleine évolution – notamment avec les règlements européens sur les services numériques, les marchés numériques et l'intelligence artificielle – qui exigent des grandes plateformes une meilleure transparence, une responsabilisation dans la

modération de contenu et une plus grande ouverture vis-à-vis de la société civile et du monde académique. Le tout dans une optique de protection des consommateurs et de respect de l'autodétermination individuelle.

Ce développement s'inscrit également dans un climat de débats de plus en plus polarisés, sur fond de rivalités idéologiques et géopolitiques croissantes. Or, l'identification n'a rien à voir avec le bien ou le mal, le vrai ou le faux: il s'agit uniquement de répondre à la question « qui est qui ». L'identité est, par définition, neutre et non-jugeante – et devrait donc être relativement épargnée par les batailles morales. Quelle que soit l'orientation politique ou les convictions personnelles, l'identification et l'attribution resteront des piliers fondamentaux de la confiance. Comme dans toute relation humaine: il faut savoir à qui l'on a affaire pour pouvoir lui faire confiance.

Ce principe s'applique aussi bien aux interactions entre humains, entreprises ou machines. Le *Global Media Identifier* apportera cette valeur ajoutée au domaine de l'information et des médias, dans l'espoir de restaurer la confiance dans un moment critique.

OLAF STEENFADT

Fondateur et Directeur Général du Global Media Registry
o.steenfadt@mediaregistry.org

[1] *Global Media Registry* est une organisation allemande à but non lucratif qui a initié et facilité le processus d'élaboration de la norme GMI par l'ISO



« alphaonetwo Crédit photo Adobe Stock – , généré à l'aide de l'IA.